

DOCUMENT RESUME

ED 115 654

TM 004 913

AUTHOR Bessent, Authella; Jennings, Earl
TITLE A Monte Carlo Study of the Analysis of Variance by Unweighted Means.
PUB DATE 75
NOTE 17p.; Paper presented at the Annual Meeting of American Educational Research Association (Washington, D.C., March 30-April 3, 1975)
EDRS PRICE MF-\$0.76 HC-\$1.58 Plus Postage
DESCRIPTORS *Analysis of Variance; Comparative Analysis; *Computer Programs; Goodness of Fit; Sampling; *Simulation; *Statistical Analysis; Tests of Significance
IDENTIFIERS Least Squares Analysis; Monte Carlo Methods; Unweighted Means Analysis

ABSTRACT

The intent of the study was to determine the extent to which test statistics computed by the unweighted means analysis are F-distributed. Applicability criteria were sought in terms of the number of factor levels and the degree to which cell frequencies differ. The unweighted means analysis, a frequently used approximate analysis, was contrasted with three least squares solutions. Evidence was relatively strong in favor of a least squares analysis if one is to conduct a two-factor analysis of variance for fixed effects. However, results confirmed that the approximate solution can be used with some confidence on main effects but not interactions when cell frequencies do not differ by more than four to one and factors exist at no more than four levels. (Author)

* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *

ED115654

A Monte Carlo Study of the Analysis of Variance
by Unweighted Means

PERMISSION TO REPRODUCE THIS COPY.
RIGHTED MATERIAL HAS BEEN GRANTED BY

Authella Bessent
TO ERIC AND ORGANIZATIONS OPERATING
UNDER AGREEMENTS WITH THE NATIONAL IN-
STITUTE OF EDUCATION. FURTHER REPRO-
DUCTION OUTSIDE THE ERIC SYSTEM RE-
QUIRES PERMISSION OF THE COPYRIGHT
OWNER.

Authella Bessent and Earl Jennings
The University of Texas at Austin

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN-
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT
OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY.

TM004 913

Abstract

The intent of the study was to determine the extent to which test statistics computed by the unweighted means analysis are F-distributed. Applicability criteria were sought in terms of the number of factor levels and the degree to which cell frequencies differ. The unweighted means analysis, a frequently used approximate analysis, was contrasted with three least squares solutions. Evidence was relatively strong in favor of a least squares analysis if one is to conduct a two-factor analysis of variance for fixed effects. However, results confirmed that the approximate solution can be used with some confidence on main effects but not interactions when cell frequencies do not differ by more than four to one and factors exist at no more than four levels.

According to Applebaum and Cramer (1974) the "nonorthogonal multifactor analysis of variance is perhaps the most misunderstood analytic technique available to the behavioral scientist, save factor analysis." There is good reason to believe that such an assertion is substantially accurate even without the qualifiers.

There appear to be multiple causes for the misunderstandings that exist. Recent work by Carlson and Timm (1974), Joe (1971), Rawlings (1972), and Ward and Jennings (1973) lead to the inference that a great deal of the confusion can be traced to texts that attempt to put in the hands of the user a set of convenient computational algorithms. The net effect of this practice has been to encourage practitioners to name their answers rather than specify their hypotheses. Thus two different individuals given the same set of data, both claiming to have performed "an unequal N's analysis of variance," may very well produce source tables with identical names for the answers but different numerical results. The obvious inference is that at least one of the individuals (possibly both) did something "wrong." In practice both may have conducted statistically defensible analyses for different hypotheses-neither explicitly stated.

Complicating a consideration of the issue is the existence of what Applebaum and Cramer call "antiquated 'approximate' methods." One such approximate method is usually called an unweighted means analysis and many currently popular texts such as Dayton (1970), Glass and Stanley (1970), Kirk (1968), and Winer (1971), cover the topic in some

detail. By careful reading in these sources, one may infer that the analysis is approximate but none of them specify clearly what it is supposed to approximate.

The origin of the method of "unweighted means analysis" can be traced to Yates (1934) who describes it as an approximate solution useful only when the class numbers do not differ very greatly. The purpose of this paper is to describe a simulation study investigating the properties of the unweighted means analysis and to contrast the results obtained with other methods based on least squares.

Description of the Simulation

Computer programs were used to sample repeatedly from populations with known distributions and to compute test-statistics for two factor designs. Resulting sampling distributions were compared with theoretical F-distributions in terms of expected values and the frequency of Type I errors. Those dimensions which were varied are given in Table 1.

Insert Table 1 about here

Definitive sources for the procedures employed for computing test statistics can be found in Table 2.

Insert Table 2 about here

Data Generation

Random samples were drawn from normally distributed populations with homogeneous variance ($\sigma^2 = 1$) about population means which ranged from 5 to 30. Each combination of the levels of variation for experimental variables was replicated one hundred times, requiring a total of $4 \times 6 \times 4 \times 100 = 9,600$ sets of data. A cell frequency pattern was generated for each of these combinations and maintained throughout the one hundred replications.

Population means were fixed according to the patterns specified in Table 1; e.g., the μ_{ij} for pattern 1, 4×3 design were as follows:

$$\begin{array}{lll} \mu_{11} = 5 & \mu_{12} = 10 & \mu_{13} = 15 \\ \mu_{21} = 15 & \mu_{22} = 10 & \mu_{23} = 5 \\ \mu_{31} = 10 & \mu_{32} = 15 & \mu_{33} = 5 \\ \mu_{41} = 10 & \mu_{42} = 5 & \mu_{43} = 15 \end{array}$$

Each set of population means remained fixed for four hundred replications—one hundred for each of the minimum/maximum cell frequency conditions.

Data Analysis

The distributions of test statistics which resulted from all four of the procedures given in Table 2 were compared with theoretical F-distributions in terms of Type I error ($\alpha = .10, \alpha = .05$, and $\alpha = .01$) and expected values. When the null hypothesis was true and both full and restricted

models were true,¹ then all assumptions for F were met; thus test statistics computed by all four methods, if accurate, should have been F-distributed.

Main effects test resulting from fitting constants and unweighted regression when the full model was not true, i.e., interaction was present, were also compared with theoretical F-distributions even though an F-distribution was not expected. Conceivably, a researcher could incorrectly assume no interaction.

Computer Programs

Even though general purpose computer programs for computing test statistics were available, new programs were written to increase efficiency. Thus cost for computer time was kept at a minimum.

Accuracy of the computer programs was tested by comparing output with that of AVAR23 (Veldman 1967) and LINEAR (Ward & Jennings, 1973). AVAR23 and LINEAR are widely used to conduct unweighted means analysis and weighted squares of means analysis, respectively. Corresponding outputs were identical. Further pattern 4 (see Table 1) was used to make a base line run with equal cell frequencies. As anticipated, unweighted means, weighted squares of means, and fitting constants analyses produced identical values which were similar to the results from the unweighted regression analysis.

Random Number Generation

Function RANF (Laurens, 1970) was used to produce pseudo-

random numbers which were subsequently used in an algorithm developed by Raiston and Wilf (1967) to obtain a random point from an $N(0, 1)$ distribution. Population means, μ_{ij} were added to produce random points from $N(\mu_{ij}, 1)$ distributions.

Summary of Results

Results from the simulation runs were examined for discrepancies between (1) the number of observed Type I errors vs. the number expected and (2) the observed mean of the calculated F's vs. the expected value. In general the observed means of the calculated F's did not differ significantly from the expected values for any of the methods. Table 3 contains evidence of the extent to which the observed frequency of Type I errors differed from the expected frequency.

Insert Table 3 about here

A marked tendency existed for the unweighted means analysis to produce more Type I errors than expected and for the least squares methods to produce fewer than expected. For example, the unweighted means interaction test at the .01 level produced almost twice (1.9028) as many errors as expected whereas the fitting constants method at the .01 level produced only 81% as many errors as expected.

Summarized in Table 4 are a series of Chi Square goodness of fit tests, comparing observed and expected frequencies at the .10, .05 and

.01 levels.

Insert Table 4 about here

The most striking results in Table 4 is the lack of fit for the unweighted means interaction test. Few trends are discernible although it should be noted that the weighted squares of means produced no significant chi squares.

Even though both fitting constants and unweighted regression are inappropriate analyses in the presence of interaction, main effects tests were conducted when interaction existed in order to determine the consequences of using these analytic procedures inadvertently. These results are not reflected in Tables 3 and 4 but neither method resulted in valid tests.

Summary

Summarizing, if one is to conduct a two-factor analysis of variance for fixed effects with unbalanced data, the evidence was relatively strong in favor of a weighted squares of means analysis, especially if a test for interaction is to be conducted. However, the approximate solution, U , can be used with some confidence to test for main effects when cell frequencies do not differ by more than four to one and factors exist at no more than four levels. Fitting constants and unweighted regression should not be used to conduct main effects tests unless the researcher is confident that interaction is negligible.

References

- Applebaum, M. I., & Cramer, E. Some Problems in the Nonorthogonal Analysis of Variance. Psychological Bulletin, 1974, 81, 335-343.
- Carlson, J. E., & Timm, N. H. Analysis of Nonorthogonal Fixed-Effects Designs. Psychological Bulletin, 1974, 81, 344-353.
- Dayton, C. M. The design of educational experiments. New York: McGraw-Hill Book Company, Inc., 1970.
- Glass, G. V., & Stanley, J. C. Statistical methods in education and psychology. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970.
- Graybill, F. A. An introduction to linear statistical models. Volume I. New York: McGraw-Hill Book Company, Inc., 1961.
- Joe, G. W. Comment on Overall and Spiegel's "Least Squares Analysis of Experimental Data." Psychological Bulletin, 1971, 75, 364-366.
- Kirk, R. E. Experimental design: procedures for the behavioral science. Belmont, California: Brooks/Cole, 1968.
- Laurens, J. (Ed.) User's Manual. Austin: Computation Center, The University of Texas at Austin, 1970, p. 13-15.
- Ralston, A., & Wilf, H. S. Mathematical methods for digital computers. Volume II. New York: John Wiley & Sons, Inc., 1967.
- Rawlings, R. R. Note on Nonorthogonal Analysis of Variance. Psychological Bulletin, 1972, 77, 373-374.
- Veldman, D. J. Fortran programming for the behavioral sciences. New York: Holt, Rinehart and Winston, 1967.

Ward, J. H., & Jennings, E. Introduction to linear models. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1973.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill Book Company, Inc., 1971.

Yates, F. The Analysis of Multiple Classifications with Unequal Numbers in the Different Classes. Journal of American Statistical Association, 1934, 29, 51-66.

Footnote

¹For a definition of a true model, see Ward and Jennings, 1973,
pp. 108-109

Table 1
Summary of Experimental Variables

Variable	Levels of Variation		
	Minimum	Maximum	
1. Degree of inequality of cell frequencies	1. 4	16	
	2. 25	100	
	3. 1	16	
	4. 16	256	
2. Number of levels in Factor A. (B Factor: held constant at 3 levels)	2, 3, 4, 5, 8, and 10 levels		
3. Patterns of population means (effects present)	A	B	A x B
	1. No	No	Yes
	2. Yes	Yes	No
	3. No	Yes	No
	4. No	No	No

Table 2
Summary of Procedures Used for Computing Test Statistics

Estimation Procedures	Definitive Source
1. Unweighted Means	Winer, 1971, pp. 402-404; 445-449 Yates, 1934
2. Weighted Squares of Means	Carlson and Timm, 1974, pp. 564-565 Yates, 1934
3. Fitting Constants (Main Effects only)	Carlson and Timm, 1974, pp. 565-566 Winer, 1971, pp. 404-414; 498-502 Yates, 1934 Kirk, 1968, pp. 204-208
4. Unweighted Regression (Main Effects only)	Carlson and Timm, 1974, p. 567: $F_A = \frac{[SS_e(\mu, \beta) - SS_e(\mu, \alpha, \beta)]/df_1}{SS_e(\mu, \alpha, \beta)/df_2}$ $F_B = \frac{[SS_e(\mu, \alpha) - SS_e(\mu, \alpha, \beta)]/df_1}{SS_e(\mu, \alpha, \beta)/df_2}$ Graybill, 1961, pp. 287-304

Table 3
 Ratio of Number of Observed
 Type I Errors to Number Expected

	α	MAIN EFFECTS	INTERACTION
Unweighted	.01	1.3472	1.9028
	.05	1.1094	1.3111
	.10	.9806	1.1486
Weighted	.01	.8472	.9167
	.05	.9389	.9500
	.10	.9264	.9972
Fitting Constants	.01	.8125	
	.05	.9208	
	.10	.9145	
Unweighted Regression	.01	.8125	
	.05	.9083	
	.10	.9167	

Table 4
Chi Square Value for the Goodness of Fit Tests

Method	2x3	3x3	4x3	6x3	8x3	10x3	4/16	25/10	1/16	16/225	
Main Effect	U	.1	4.7	1.8	10.4**	1.5	6.8*	1.1	10.8**	5.5	17.1***
	W	.1	.2	4.9	4.6	5.1	2.7	1.2	4.9	1.1	2.6
	FC	.6	.8	.9	.4	1.5	7.6*	6.2*	13.8*	.5	1.5
	R	.4	.8	1.9	.3	1.8	9.3**	5.6	13.8**	.0	1.5
Interaction	U	1.0	6.5*	8.3*	16.6**	41.3**	19.5**	12.2**	3.3	42.6**	67.8**
	W	2.9	3.8	4.7	.1	3.9	2.4	3.0	3.7	4.4	1.4
* P .05 ** P .01 with 2 df											

* P .05

** P .01 with 2 df